Running a Correspondence Analysis (CA) from a contingency table with XLSTAT

demoCA.xls

Dataset for running a Correspondence Analysis

An Excel sheet with both the data and the results can be downloaded by clicking here.

The data correspond to a survey asking moviegoers their opinion on a film they had just seen. The audience was also asked to give their age category.

Setting up a Correspondence Analysis

Once XLSTAT is open, select the XLSTAT / Analyzing data / Correspondence analysis command, or click on the corresponding button of the Analyzing Data toolbar (see below).

Once you have clicked on the button, the Correspondence analysis dialog box appears.

Select the data on the Excel sheet. If your data are in a pivot table format (as in the example below), select the Contingency table format.
Note: If your data are in an Observations/variables format, select the corresponding option.

As the names of the categories of the contingency table were included, the Labels included option was selected as well.

We selected the Range option for the output. The $J$4 cell was selected as the upper left corner of the results report.
You can also choose to write the results in a separate sheet or workbook.

On the charts tab, we selected to display the symmetric rows and columns plot.
Interpretation of the Correspondence Analysis

The results are displayed once the user has selected and validated the axes on which the plots need to be displayed.

A Chi-square test is computed to test if the rows and columns are independent. As the p-value is lower than the significance level (0.05) we conclude that the rows and columns are not independent, which means there is relationship between the rows and the columns.

The quality of the analysis can be evaluated by consulting the table of the eigenvalues or the corresponding scree plot. If the sum of the two (or a few) first eigenvalues is close to the total represented, then the quality of the analysis is very high. The correspondence analysis in this example is of good quality as the sum of the first two eigenvalues adds up to 97% of the total.

The most interesting result in Correspondence analysis is the map of the categories including both rows and columns. If the quality of the analysis is good (97% as in this example), the map
can be used to interpret the data. We can see here that among the people queried, those between the ages of 16 and 24 were more inclined to like the film than those belonging to other age categories. Those in the 55-64 age category were, as a group, the least favorable towards it.

Correspondence analysis is a very effective technique for analyzing 2-way tables. When more than two categorical variables are used in a survey, the best technique to use is Multiple Correspondence analysis (MCA).

Watch this video to see this tutorial demonstrated.

http://www.youtube.com/watch?feature=player_embedded&v=O6RyRhczZD8

**Setting up a Correspondence Analysis**

Once XLSTAT is open, select the XLSTAT / Analyzing data / Correspondence analysis command, or click on the corresponding button of the Analyzing Data toolbar (see below).
Once you have clicked on the button, the Correspondence analysis dialog box appears.

Select the data on the Excel sheet. If your data are in a pivot table format (as in the example below), select the **Contingency table** format.

*Note: If your data are in an **Observations/variables** format, select the corresponding option.*

As the names of the categories of the contingency table were included, the **Labels included** option was selected as well.

We selected the **Range** option for the output. The $J$4 cell was selected as the upper left corner of the results report.

You can also choose to write the results in a separate sheet or workbook.
On the **charts** tab, we selected to display the **symmetric rows** and **columns plot**.
Interpretation of the Correspondence Analysis

The results are displayed once the user has selected and validated the axes on which the plots need to be displayed.

An Chi-square test is computed to test if the rows and columns are independent. As the p-value is lower than the significance level (0.05) we conclude that the rows and columns are not independent, which means there is relationship between the rows and the columns.

The quality of the analysis can be evaluated by consulting the table of the eigenvalues or the corresponding scree plot. If the sum of the two (or a few) first eigenvalues is close to the total represented, then the quality of the analysis is very high. The correspondence analysis in this example is of good quality as the sum of the first two eigenvalues adds up to 97% of the total.

The most interesting result in Correspondence analysis is the map of the categories including both rows and columns. If the quality of the analysis is good (97% as in this example), the map can be used to interpret the data. We can see here that among the people queried, those between the ages of 16 and 24 were more inclined to like the film than those belonging to other age categories. Those in the 55-64 age category were, as a group, the least favorable towards it.
Correspondence analysis is a very effective technique for analyzing 2-way tables. When more than two categorical variables are used in a survey, the best technique to use is Multiple Correspondence analysis (MCA).

Watch this video to see this tutorial demonstrated.

http://www.youtube.com/watch?feature=player_embedded&v=O6RyRhczZD8